

CS5512: Machine Learning

Programming Assignment 1

Due on 10 Feb

For programming assignment please submit your code and a short discussion on your observation (preferably PDF and latex) from your experiments. Put all codes and report in a single zipped file and name it as <First-name><Last-name>.zip. Then submit it in moodle.

1 Programming Exercises

Exercise 1 : File read and write [point 1]

(1 point) Download iris data from <https://archive.ics.uci.edu/ml/machine-learning-databases/iris/>. This is perhaps the best known database to be found in the pattern recognition literature. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant (Iris Setosa, Iris Versicolour, Iris Virginica). It contains 4 numeric features and the predicted attribute is class of iris plant.

Write a program (without using any machine learning related in built library) to create another file named as *iris-sum-input.txt* which contain the same data but in the following format

- Each line of the output file contains information corresponding to one sample in the following format

label index1:value1 index2:value2

- *index* indicates the index of feature and *value* indicates the numerical value of that feature for the sample.
- Skip the *index* if the corresponding *value* is zero
- Consider class labels as follows
 - Iris-setosa : 1
 - Iris-versicolor : 2
 - Iris-virginica : 3

- example : class 1, the feature vector (0.7,1,1,0,2) translates to

1 1:0.7 2:1 3:1 5:2

1.1 Exercise 2 : Regression [point 3+1]

In the following you need to learn a function, $f : \mathbb{R} \rightarrow \mathbb{R}$, given some training data $x \in \mathbb{R}, y \in \mathbb{R}$ using regression.

Data is in *assignment1.zip*. It contains training data file with

- x : training feature
- Y : training label

and test data file with

- x_{ts} : test feature

The relation between x and y can be non-linear. Hence to catch non-linear relationship we will generate Feature matrix= $[1, \mathbf{x}, \mathbf{x}^2, \dots, \mathbf{x}^{10}]$

You need to submit following:

- (point 2) code in python(.py) or matlab(m). [Please do not submit notebook file.]
- code should input training data and test data from file as given write prediction in submission.csv with two column.
- (point 1) Write a report by given details about loss function, regularization and validation method you have used and why? Discuss with increase of regularization parameter how the training error and validation error change.
- (point 1) In case your submission file is proper I shall calculate your prediction loss and this point will be assigned to best 10 predictor.